BEILSTEIN INSTITUT

Discovering the Subtleties of Sugars
June 10th – 14th, 2013, Potsdam, Germany

UniCarbKB: First Year Report Card

Matthew P. Campbell[1], Robyn Peterson[1], Elisabeth Gasteiger[2], Jingyu Zhang[1], Yukie Akune[3], Jodie L. Abrahams[1], Julien Mariethoz[4], Catherine A. Hayes[5], Daniel Kolarich[6], Niclas G. Karlsson[5], Kiyoko F. Aoki-Kinoshita[3], Frédérique Lisacek[4], Nicolle H. Packer[1,*]

[1]Biomolecular Frontiers Research Centre, Macquarie University, North Ryde, Sydney, NSW 2109, Australia,

[2]Swiss-Prot Group, Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland

[3]Department of Bioinformatics, Faculty of Engineering, Soka University, 1 – 236 Tangi-machi, Hachioji, Tokyo, Japan

[4]Proteome Informatics Group, Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland

[5]Department of Medical Biochemistry and Cell Biology, University of Gothenburg, S-40 530 Gothenburg, Sweden

[6]Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

E-Mail: *nicki.packer@mq.edu.au

Received: 19th December 2013 / Published: 22nd December 2014

Abstract

At the Beilstein Workshop on Glycoinformatics in 2011 we introduced UniCarbKB as an international initiative that aims to collect and distribute resources and practices from glycobiology practitioners to the whole biological research community. The mission was, and still is, to provide a comprehensive, high quality catalogue of information on carbohydrates, and to continue efforts to advance the interpretation of captured data through the development of novel data analysis methods and algorithms for the efficient representation and mining of large experimental data sets.

This article is part of the Proceedings of the Beilstein Glyco-Bioinformatics Symposium 2013.
www.proceedings.beilstein-symposia.org

Here, we report the progress we have made on establishing the infrastructure and content of the fledgling UniCarbKB. This will include our current work on the integration, into the publically available UniCarbKB portal, of data from UniCarb-DB, GlycoSuiteDB, GlycoBase, EUROCarbDB, SugarBind and PubChem. In the future it is hoped that the UniCarbKB knowledgebase, based on a central database of curated glycan structures, will become the key resource of quality information for glycobiology research.

Introduction

The NIH report Transforming Glycoscience: A Roadmap for the Future [1] identified that an important factor in extending the outreach of glycomics is the urgent need for databases to store, process and disseminate structural and analytical datasets. The widely acknowledged sparseness of maintained resources continues to hamper the realisation tools, which are increasingly required to support high-throughput glycomics. During the 2nd Beilstein Glyco-Bioinformatics Symposium 2011 members of the UniCarbKB consortium presented a long-term vision [2] to build an infrastructure that adopts and extends the principles of quality shared by GlycoSuiteDB [3] and EUROCarbDB [4]. This new open-access infrastructure (called UniCarbKB) shall constitute the nucleus for a central depository for carbohydrate-related data (structure and function), comparable to and cross-linked with the extensively used genomics and proteomics data collections.

UniCarbKB: Laying the Foundations in Year One

UniCarbKB strives to radically enhance the infrastructure required to better enable glycomics research by making data more accessible, and presenting a single, user-friendly interface to a growing range of resources being developed. UniCarbKB was officially launched at the 3rd Beilstein Glyco-Bioinformatics Symposium 2013 [5]. The launch highlighted our mission to support glycomics and glycobiology research by providing a comprehensive, richly and accurately annotated glycan structure knowledgebase, with extensive cross-references and a redesigned intuitive user interface. The project is maturing and will continue to do so, but even at this early stage it has started to address major concerns raised by the research community and issues outlined in the 2012 NIH report.

For year one our design ethos has focused on providing researchers with 1) long-term and open-access to a highly-curated database of experimentally determined glycan structures; by 2) establishing a centralised model for merging curated data collections from GlycoSuiteDB, GlycoBase [6], UniCarb-DB [7] and EUROCarbDB; and 3) increasing awareness of data collections by connecting UniCarbKB with other databases such as SugarBind, a database of pathogen–sugar interactions [8]. Such activities include the refactoring of GlycoSuiteDB and the modification of modules from EUROCarbDB necessary to support

existing and future structure and analytical data collections. During the engineering phase the developers systematically explored the functionality and database designs of EUROCarbDB and GlycoSuiteDB. Here, the database structure integrates components from previous efforts to provide a more flexible and comprehensive relational database. It is likely that the schema will continually evolve with new data collections and data requirements. For example, core features of EUROCarbDB have been retained including GlycanBuilder [9, 10], MonosaccharideDB (http://www.monosaccharidedb.org) and the encoding standards developed in the context of GlycoCT [11].

Early phase developments are focused on enhancing existing tools, standards and applications to be more accessible and amenable to modern research workflows. In particular we have leveraged previous experiences to build a modern and scalable framework, which uses technologies and web frameworks that are more familiar to developers.

Establishment of the eResearch Infrastructure

Support from the Australian National eResearch Collaboration Tools and Resources (NeCTAR) provides UniCarbKB and its affiliated activities access to a sustainable infrastructure. A core principle of NeCTAR is to manage and preserve valuable data collections through the provision of robust data service architecture. More importantly the significant data centric investments will be maintained and shall form a critical component of life science research, thus ensuring that data will not disappear. In addition, the longevity of the resource shall be optimised by the mirroring of UniCarbKB on the ExPASy server [12], and by establishing close ties with the UniProt protein knowledgebase [13]. In the initial scope of the project the partners are establishing a framework of connected services comprising 1) a dedicated web-application front-end hosting the searchable main glycan structural database and 2) a mass spectrometry data collection supported by the efforts of UniCarb-DB and 3) the development of processing and interpretation tools.

Designing the UniCarbKB Experience

We have worked hard to put user experience design at the heart of the development cycle. When starting UniCarbKB the developers interacted with researchers and identified a series of ‘stories’ or specific themes that would need to be supported. It was critical, at an early stage, to identify the requirements and expectations of the community; for example we asked what features of EUROCarbDB and GlycoSuiteDB users frequently used and if/what improvements need to be considered. A series of design sprints and wire-framing sessions were conducted between the developers and the identified user base. Design sprints included a series of prototypes and iterative usability-based refinements of the design, evaluation of page layout, and generation of wireframes. By adopting agile development practices the developers and designers were able to build an application that is easy to use and allows researchers to accomplish their goals. This is being continually tested by a core group of users for feedback and improvements.

A detailed description of the new user-interface is documented in the *Nucleic Acids Research Database Issue* [5]. In brief, the new interface is more visual, encapsulating a simpler content layout with an emphasis placed on displaying information that researchers want to access. Many of these changes are in line with the presentation of information previously available from EUROCarbDB and GlycoSuiteDB databases and a choice of symbol nomenclatures [14, 15]. The goal is to retain the established user-base and minimise distraction with the launch of UniCarbKB. The web-application user experience is built using Twitter Bootstrap that provides designers access to rich features namely: responsive design, Javascript interaction, typography and cross-browser compatibility. It provides a faster, easier and less repetitive solution for delivering a robust front-end. By making use of these features plus the integration of JavaScript libraries (notably jQuery) UniCarbKB is delivering an enhanced user experience.

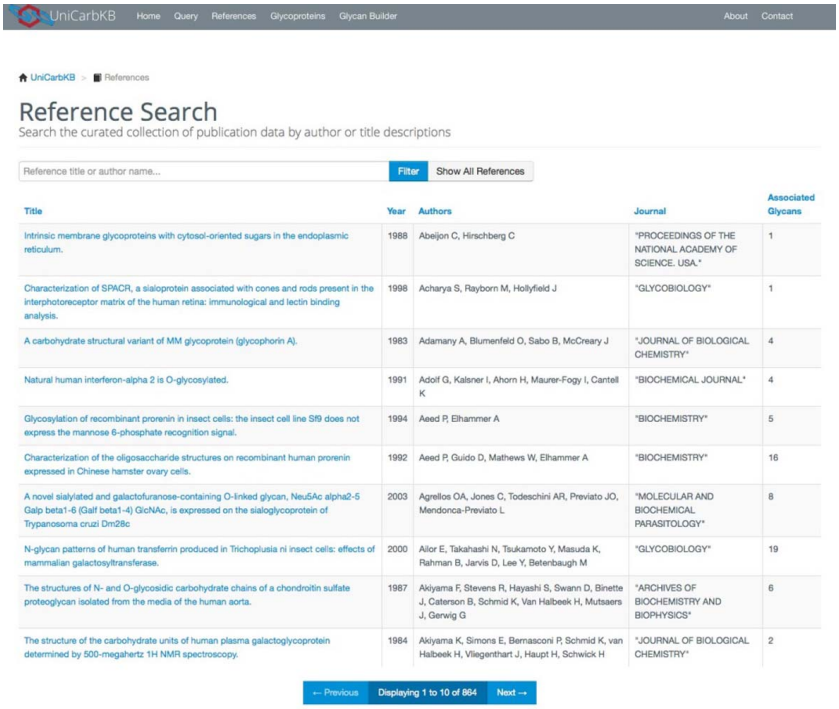

Figure 1. UniCarbKB has a completely revamped user-interface that focuses on presenting information in a clean and concise manner. For example, when browsing the reference collections users can search for content using the improved 'Filter' bar; searches can include author name or publication keywords. Pagination is used throughout UniCarbKB to improve both data layout and content navigation. The design of UniCarbKB was achieved by actively working alongside researchers to provide an intuitive data rich experience.

For example, to simplify searching and to organise result outputs we have made better use of auto completion, pagination and managed the efficient display of associated data links. Figure 1 shows how the user can readily search the curated publication information and pagination allows for better handling of large result sets. Figure 2 shows the summary page for the protein alpha-1-acid glycoprotein, which provides a description of the attached glycan structures and knowledge of site-specific glycosylation that has been curated from the literature. In addition, for each protein summary we provide a comprehensive summary of associated metadata including biological source and publications citing the data.

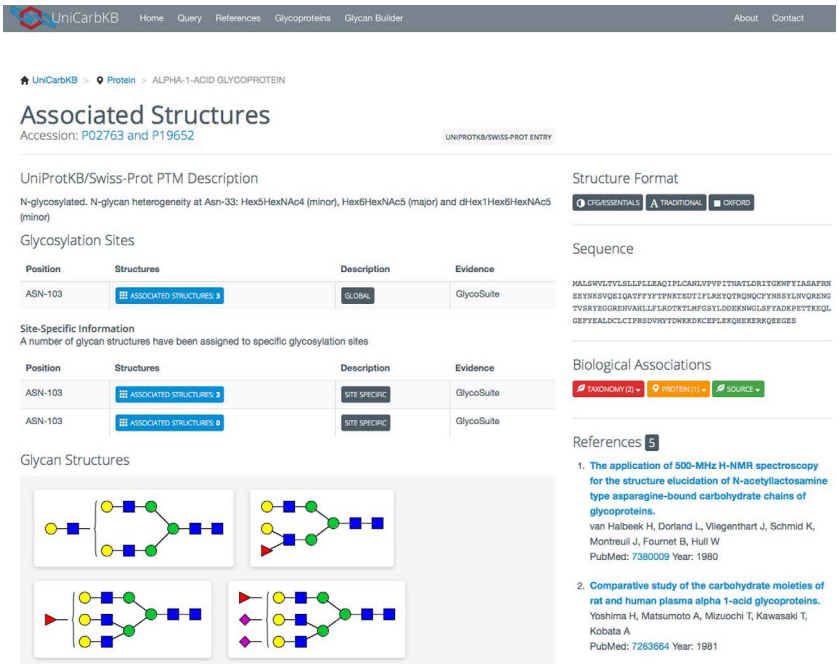

Figure 2. An example protein summary page for alpha-1-acid glycoprotein. The entry provides a description of the glycan structures characterised on this glycoprotein and the number of structures associated with experimentally confirmed glycosylation sites. Information is provided on the 'Biological Associations' including details on species and tissue source in addition to the protein sequence extracted from UniProtKB. The 'References' that have been used to curate the information are summarised in the right side panel including PubMed links.

UniCarbKB Biocuration

At this stage in the project UniCarbKB's central focus is the manual annotation of structural information from the literature, limited to mammalian *N*- and *O*-linked glycan structures. Journal articles provide the main source of experimental knowledge, with the full text of each article being read and the information on glycan composition, sequence, linkage, tissue

and species source, attached protein if known, and the methods used to determine these characteristics being extracted manually. The aim of this approach is to build a central hub for glycan structures with an emphasis on quality glycoprotein information at the global and site-specific level. This has the advantage of providing a gold standard set of structural data in the knowledgebase, which can be used to build pipelines for the automatic capture of related information.

With the improvements in mass spectrometric technology, UniCarbKB is expanding from the initial focus of GlycoSuiteDB and EUROCarbDB in only collecting global glycoprotein glycan structures as described in the literature to now include site-specific glycosylation when known. For example, UniCarbKB curators are mining a literature review of 117 research papers encompassing over 400 glycosylation sites from more than 160 mammalian *N*-glycoproteins [16]. On-going efforts by the small curation team to include these data collections in the UniCarbKB knowledgebase will add considerable value to the existing 400 glycoproteins and 598 glycosylation sites now available. This site-specific glycosylation information will be linked to the relevant glycoprotein entries in UniProtKB.

Web Services and Semantic Technologies

Data-sharing and open-access models are increasingly important in the dissemination of knowledge in the life sciences. Increasingly, developers and users of bioinformatic platforms require access to well-documented libraries to create applications that query and make use of data collections available in public resources. To this end UniCarbKB supports programmatic access via web services to further facilitate the reuse of curated data collections across tools, other databases and research disciplines.

The web services build upon the efforts initiated by the Working Group for Glycomics Data(base) Standards (WGGDS) supported by the Consortium of Functional Glycomics (CFG) in 2010. In brief, the WGGDS was tasked with designing a cross-database interface based on RESTful protocols for querying and retrieving content from affiliated databases. To support the philosophy of UniCarbKB by providing the community with access to a high-quality database infrastructure the team have reengineered and extended components of the original protocols. UniCarbKB-WS is being developed with the JBossWS framework using enterprise standard protocols to provide the community with a first-of-its-kind secure and readily accessible remote querying resource.

A beta-release of UniCarbKB-WS is in the final stage of user testing prior to public release. Access to this resource, hosted on Australian NeCTAR infrastructure and mirrored on the Swiss ExPASY servers, will allow users to query the database via (sub)structure searching, structure IDs and glycan mass. A detailed description of the services available to developers is provided at http://dev.unicarbkb.org with an envisaged launch in early 2014.

Although the web services and query features on the website are efficient search tools there is an increasing demand in the life sciences for the provision of semantic web technologies. This is exemplified by the launch of the EMBL-EBI RDF platform (October 2013) that adopts the SPARQL Protocol and RDF Query Language, which provides a unified mechanism to query across multiple resources. In order to better integrate glycomics data collections, a glycoRDF working group comprising glycodatabase developers from Japan, Australia, Russia, Germany and USA has been established [17] and a standardised RDF document that is representative of all major glycan databases is being formalised. The access to a unified model will forge connections between existing resources that will considerably enhance data discovery and allow researchers to ask more complex biological questions. To this end, UniCarbKB developers are now contributing towards the generation of a compatible RDF framework, in agreement with members of the project, that includes (but is not limited to) a dedicated SPARQL endpoint and access to a regularly updated triplestore.

Connecting UniCarbKB with UniProtKB

There have been few intensive programmes to cross-reference glycan related databases with those that support genomics and proteomics research. Previously, GlycoSuiteDB was the sole provider of curated glycoprotein data to UniProtKB [12], however, during the last year members from both database activities have commenced a programme to share new glycoprotein knowledge through agreed formats. This collaborative effort, although in its infancy, aims to establish a mechanism whereby new content gained by UniCarbKB is integrated into UniProtKB on a regular basis.

A milestone towards this long-term agreement was reached in May 2013 with UniProtKB switching to UniCarbKB links instead of (the no longer maintained) GlycoSuiteDB (UniProt release 2013_06). Subsequently, in the UniProt release 2013_08 the GlycoSuiteDB name was substituted by UniCarbKB; such that all UniProtKB glycosylation entries associated with GlycoSuiteDB have now been updated to link with UniCarbKB and *vice-versa*.

To ensure accurate cross-referencing the developers standardised and defined ontology terms to link the two databases. Updates to relevant UniProtKB records and the inclusion of new unique identifiers (IDs) to UniCarbKB glycan structures allow users direct access to structural information and corresponding meta-data for relevant glycoproteins. In agreement with senior developers at UniProtKB new glycoprotein data curated by UniCarbKB will be integrated into UniProtKB and information flow will be two-way between the databases.

GlycoMod: Linking Monosaccharide Composition with Glycan Structures

GlycoMod [18] is a well-established programme designed to determine possible glycan structure compositions from experimentally determined mass. The tool is part of the proteomics suite of tools available on ExPASy. Indeed, it was an innovative tool that in 2001 recognised the importance of connecting proteomics and glycomics. The programme can be used to predict the composition of a glycan comprised of either underivatised, methylated or acetylated monosaccharides, or with a derivatised reducing terminus. Furthermore, the composition of a glycan attached to a glycopeptide can be calculated if the sequence or mass of the attached peptide is known. In such instances GlycoMod communicates with the UniProt Knowledgebase databases and matches experimentally determined glycopeptide masses against predicted protease-produced peptides, which enables the prediction of the composition of *N*- and *O*-linked oligosaccharides on glycopeptides.

We have directly linked the GlycoMod tool with UniCarbKB. Here, the two database initiatives actively share compositional and glycan structural data respectively. For each theoretical composition calculated by GlycoMod links to UniCarbKB will provide users with a list of composition-structure matches reported in the literature.

Integrating N-glycan Biosynthesis Pathways to Help Validate Structure Entries

Recently, we have initiated a project to integrate information about known genes and enzymes involved in the biosynthesis of *N*-glycans. This new platform, called GlycanSynth, involves the comprehensive curation of data related to enzyme activity, which is primarily sourced from the Encyclopedia of Genes and Genomes (KEGG) [19] and GlycoGene [20] databases. Additional information is also being gathered from the Consortium for Functional Glycomics (CFG) [21], Carbohydrate-Active enzymes (CAZy) [22, 23], BRENDA [24] and UniProt databases.

From this acquired information we have constructed a set of disaccharide reactions that match each *N*-glycosylation-related gene against donor and acceptor substrate. By using these reaction rules it will be possible for us to (i) connect gene function with glycan structure and (ii) validate the accuracy of structures stored in UniCarbKB based on acquired knowledge of the glycosylation machinery.

Forging New Connections with PubChem

PubChem (http://pubchem.ncbi.nlm.nih.gov) [25] is a public repository of molecules and their biological properties, containing more than 25 million chemical structures and 90 million bioactivity outcomes. It is the mission of PubChem to deliver free and easy access to deposited data, and to provide intuitive data analysis tools. The platform consists of three interconnected databases: Substance, Compound and BioAssay. It is the Substance database that contains detailed descriptions of molecules provided by depositors; the Compound database consists of unique chemical structures derived by structural standardisation of the records in the Substance section; while the BioAssay provides screening results of substances by assay providers.

In April 2013 PubChem and UniCarbKB announced the first phase of a programme to share glycan structural information. The aim of the programme is to increase the number and quality of biologically relevant protein-attached glycans available to researchers in PubChem. Stage one focused on validating PubChem's workflow for handling and processing a subset of fully defined UniCarbKB structures encoded in the IUPAC format. PubChem has been collaborating with NextMove on the development of 'Sugar & Splice' to convert representation of glycans. This was the first large activity that required the conversion of IUPAC notation to the SMILES representation. This collaboration will continue to grow with UniCarbKB providing regular updates to PubChem in parallel with future curation plans.

Conclusion

In 2011 we announced plans to implement a glycosciences knowledge platform that would set the state- of-the-art foundation infrastructure for the free sharing and dissemination of data on glycoconjugates, and which would seed the development of enhanced glyco-informatic tools with which to interpret the data. By 2013, in collaboration with international partners and with limited resources, we have made significant strides toward meeting our ambitious goals in a short time span.

Such efforts have spanned the development of a new web-application framework with an entirely new approach to user-interface designs and the introduction of new technologies including RDF Semantics. Considerable advancements have afforded 1) an amalgamated glycan structure database that brings together collections curated and sourced from legacy efforts; 2) a technical framework that is now enabling the consortia to deploy tools and workflows to assist the interpretation of experimental (principally mass spectrometry at this stage) data; 3) a biocuration programme to capture newly published data to significantly extend the availability of high-quality datasets and 4) established crosslinks with the proteomics knowledgebase UniProtKB and the chemical repository PubChem. Our biocuration efforts have also provided a forum for us to open discussions with Thomson

Reuters Data Citation Index (DCI) to adopt UniCarbKB as the prime indexing site for glycomics. This activity stems from our work with the Australian National Data Services (ANDS) to promote the discovery and re-use of research data. It is envisaged that early in year two DCI-ANDS will harvest data directly from UniCarbKB to link the resources directly; offering users an enhanced data search platform.

UniCarbKB is a user driven resource and feedback is extremely valuable to help us improve the content of our databases and the services offered in terms of accuracy and usability. To promote active outreach we have released a feedback tool (accessible on all pages of UniCarbKB) for users to provide comments. We are keen to work with researchers to include newly published data or updates, and as such we strongly encourage the use the forms available at http://www.unicarbkb.org/contribute to engage with our small data curation team. New releases will be published every three months accompanied by a description of changes and inclusions.

UniCarbKB has been acknowledged as a necessary, valuable and quality resource for the glycoscience community. Strong foundations for the expansion of the knowledgebase have now been set but require international buy-in with real resources for its continued development.

Acknowledgements

The authors are grateful for funding support from: The Australian National eResearch Collaboration Tools and Resources project [NeCTAR RT016 to M.P.C and N.H.P]; Australian National Data Service (ANDS) through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative to M.P.C and N.H.P]; Swiss National Science Foundation [SNSF 31003A_141215 J.M.]; Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI [F.L. and E.G.]; ExPASy is maintained by the web team of the Swiss Institute of Bioinformatics and hosted at the Vital-IT Competency Center; UniCarbKB was also supported by Agilent's University Relations programme to M.P.C and N.H.P; GlycoSuiteDB was developed by Proteome Systems Ltd [N.H.P] and transferred to SIB in 2009; The Swedish Foundation for International Cooperation in Research and Higher Education to M.P.C, C.A.H and N.G.K.

References

[1] Transforming Glycoscience: A Roadmap for the Future. 2012, National Research Council.: Washington (DC).

[2] Campbell, M.P., Hayes, C.A., Struwe, W.B., Wilkins, M.R., Aoki-Kinoshita, K.F., Harvey, D.J., Rudd, P.M., Kolarich, D., Lisacek, F., Karlsson, N.G., Packer, N.H. (2011) UniCarbKB: Putting the pieces together for glycomics research. *Proteomics* **11**:4117 – 4121.
doi: http://dx.doi.org/10.1002/pmic.201100302.

[3] Cooper, C.A., Harrison, M.J., Wilkins, M.R., Packer, N.H. (2001) GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Research* **29**:332 – 335.
doi: http://dx.doi.org/10.1093/nar/29.1.332.

[4] von der Lieth, C.-W., Freire, A.A., Blank, D., Campbell, M.P., Ceroni, A., Damerell, D.R., Dell, A., Dwek, R.A., Ernst, B., Fogh, R., Frank, M., Geyer, H., Geyer, R., Harrison, M.J., Henrick, K., Herget, S., Hull, W E., Ionides, J., Joshi, H.J., Kamerling, J.P., Leeflang, B.R., Lütteke, T., Lundborg, M., Maass, K., Merry, A., Ranzinger, R., Rosen, J., Royle, L., Rudd, P.M., Schloissnig, S., Stenutz, R., Vranken, W.F., Widmalm, G., Haslam, S.M. (2011) EUROCarbDB: An open-access platform for glycoinformatics. *Glycobiology* **21**:493 – 502.
doi: http://dx.doi.org/10.1093/glycob/cwq188.

[5] Campbell, M.P., Peterson, R., Mariethoz, J., Gasteiger, E., Akune, Y., Aoki-Kinoshita, K. F., Lisacek, F., Packer, N.H. (2013) UniCarbKB: building a knowledge platform for glycoproteomics. *Nucleic Acids Research* **42**:D215 –D221.
doi: http://dx.doi.org/10.1093/nar/gkt1128.

[6] Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A., Rudd, P.M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis *Bioinformatics* **24**(9):1214 – 1216.
doi: http://dx.doi.org/10.1093/bioinformatics/btn090.

[7] Hayes, C.A., Karlsson, N.G., Struwe, W.B., Lisacek, F., Rudd, P.M., Packer, N.H., Campbell, M.P. (2011) UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics* **27**:1343 – 1344.
doi: http://dx.doi.org/10.1093/bioinformatics/btr137.

[8] Shakhsheer, B., Anderson, M., Khatib, K., Tadoori, L., Joshi, L., Lisacek, F., Hirschmann, L, Mullen, E. (2013) SugarBind database (SugarBindDB): a resource of pathogen lectins and corresponding glycan targets. *Journal of Molecular Recognition* **26**(9):426 – 431.
doi: http://dx.doi.org/10.1002/jmr.2285.

[9] Ceroni, A., Dell, A., Haslam, S.M. (2007) The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code for Biology and Medicine* **2**:3.
doi: http://dx.doi.org/10.1186/1751-0473-2-3.

[10] Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., Dell, A., Haslam, S.M. (2012) The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biological Chemistry* **393**(11):1357 – 1362.
doi: http://dx.doi.org/10.1515/hsz-2012-0135.

[11] Herget, S., Ranzinger, R., Maass, K., von der Lieth, C.-W. (2008) GlycoCT – a unifying sequence format for carbohydrates. *Carbohydrate Research* **343**(12): 2162 – 2171.
doi: http://dx.doi.org/10.1016/j.carres.2008.03.011.

[12] Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, E., Moretti, S., Mostaguir, K., Redaschi, N., Grégoire Rossier, G., Xenarios, I., Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* **40**(Web Server issue): W597 –W603.
doi: http://dx.doi.org/10.1093/nar/gks400.

[13] The UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Research* 2013. **41**(Database issue): D43 –D47.
doi: http://dx.doi.org/10.1093/nar/gks1068.

[14] Varki, A., Cummings, R.D., Esko, J.D., Hudson H. Freeze, H.H., Stanley, P., Marth, J.D., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (2009) Symbol nomenclature for glycan representation. *Proteomics* **9**(24):5398 – 5399.
doi: http://dx.doi.org/10.1002/pmic.200900708.

[15] Harvey, D.J., Merry, A.H., Royle, L., Campbell, M.P., Dwek, R.A., Rudd, P.M. (2009) Proposal for a standard system for drawing structural diagrams of *N*- and *O*-linked carbohydrates and related compounds. *Proteomics* **9**(15):3796 – 3801.
doi: http://dx.doi.org/10.1002/pmic.200900096.

[16] Thaysen-Andersen, M., Packer, N.H. (2012) Site-specific glycoproteomics confirms that protein structure dictates formation of *N*-glycan type, core fucosylation and branching. *Glycobiology* **22**(11):1440 – 1452.
doi: http://dx.doi.org/10.1093/glycob/cws110.

[17] Aoki-Kinoshita, K.F., Bolleman, J., Campbell, M.P., Kawano, S., Kim, J.-D., Lütteke, T., Matsubara, M., Okuda, S., Ranzinger, R., Sawaki, H., Shikanai, T., Shinmachi, D., Suzuki, Y., Toukach, P., Yamada, I., Packer, N.H., Narimatsu, H. (2013) Introducing glycomics data into the Semantic Web. *Journal of Biomedical Semantics* **4**(1):39.
doi: http://dx.doi.org/10.1186/2041-1480-4-39.

[18] Cooper, C.A., Gasteiger, E., Packer, N.H. (2001) GlycoMod – a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**(2):340 – 349.
doi: http://dx.doi.org/10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B.

[19] Kanehisa, M., Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**(1):27 – 30.
doi: http://dx.doi.org/10.1093/nar/28.1.27.

[20] Narimatsu, H. (2004) Construction of a human glycogene library and comprehensive functional analysis. *Glycoconjugate Journal* **21**(1 – 2):17 – 24.
doi: http://dx.doi.org/10.1023/B:GLYC.0000043742.99482.01.

[21] Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S., Sasisekharan, R. (2006) Advancing glycomics: implementation strategies at the consortium for functional glycomics. *Glycobiology* **16**(5):82R– 90R.
doi: http://dx.doi.org/10.1093/glycob/cwj080.

[22] Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., Henrissat, B. (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**(D 1):D 490 –D 495.
doi: http://dx.doi.org/10.1093/nar/gkt1178.

[23] Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**(Database issue):D 233 –D 238.
doi: http://dx.doi.org/10.1093/nar/gkn663.

[24] Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., Schomburg, D. (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Research* **41**(Database issue):D 764 –D 772.
doi: http://dx.doi.org/10.1093/nar/gks1049.

[25] Bolton, E.E., Wang, Y., Thiessen, P.A., Bryant, S.H. (2018) PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annual Reports in Computational Chemistry* **4**:217 – 241.
doi: http://dx.doi.org/10.1016/S1574-1400(08)00012-1